Wrinkle Detection using Hessian Line Tracking

Choon-Ching Ng, *Member, IEEE*, Moi Hoon Yap, *Member, IEEE*, Nicholas Costen, Baihua Li

***Abstract*—Wrinkles play an important role in face-based analysis. They have been widely used in applications such as facial retouching, facial expression recognition and face age estimation. Although a few techniques for wrinkle analysis have been explored in the literature, poor detection limits the accuracy and reliability of wrinkle segmentation. Therefore, an automated wrinkle detection method is crucial to maintain consistency and reduce human error. In this paper, we propose Hessian Line Tracking (HLT) to overcome the detection problem. HLT is composed of Hessian seeding and directional line tracking. It is an extension of a Hessian filter; however it significantly increases the accuracy of wrinkle localization when compared with existing methods. In the experimental phase, three coders were instructed to annotate wrinkles manually. To assess the manual annotation, both intra- and inter-reliability were measured, with an accuracy of 94% or above. Experimental results show that the proposed method is capable of tracking hidden pixels; thus it increases connectivity of detection between wrinkles, allowing some fine wrinkles to be detected. In comparison to the state-of-the-art methods such as the CUla Method (CUM), FRangi Filter (FRF), and Hybrid Hessian Filter (HHF), the proposed HLT yields better results, with an accuracy of 84%. This work demonstrates that HLT is a remarkably strong detector of forehead wrinkles in 2D images.**

***Index Terms*—Wrinkle detection, Hessian filter, line tracking, Bosphorus dataset, Jaccard similarity index**

I. Introduction

The wrinkle is a small furrow or crease in the skin, especially of the face, caused by aging [1] or expression [2]. In 2D skin images, the process of wrinkling often creates deep creases and causes curvature in the surrounding skin. The resulting skin curvature causes specific intensity gradients in images which are perceived as discontinuities and arbitrary patterns in surrounding skin textures [3].

The quantitative assessment of skin condition has been an area of quite intense study. There is great interest in supplementing the dermatologist's diagnostic visual assessment of skin with objective measures. These techniques are also valuable for the efficient development of effective pharmaceutical treatments [4]. Many skin assessment methods have been developed over the past few years. For example, analysis of the skin surface around pores on the face [5], evaluation of facial wrinkles' development over lifetime [6], and assessment of facial wrinkles using automatic detection and quantification methods [7]. Most of these studies were based on clinical judgements (subjective assessment) rather than computer vision (objective assessment). Clinicians' perspectives are typically focussed on the level of wrinkle severity, which is assessed using either descriptive scales or is photographically-calibrated [8]. In contrast, in computer vision, the concern is how precisely a wrinkle is located relative to the ground truth. The well-known shortcomings of subjective assessment, such as human errors, inconsistency of annotation and the large amount of time consumed by it limit the scientific study of the treatment and progression of skin aging. Therefore, this work highlights how to detect wrinkles accurately by assessing how closely a wrinkle is located through the Jaccard Similarity Index (JSI).

The automated detection of wrinkles from 2D images is an important step in age estimation [9] and age synthesis [10], [11]. Aznar-Casanova et al. [1] investigated the influence of wrinkles on facial age judgement. Their results indicate that the number of wrinkles and the depth of furrows are highly correlated with the perceived facial age. Ng et al. [12] studied a local wrinkle-based extractor for age estimation. They used a Canny edge detector to detect the wrinkles and presented it as a pattern of age estimation. However, an edge detector is not suitable for wrinkle detection because it only detects the boundaries. Huang et al. [2] presented a robust facial expression recognition method involving skin wrinkles. Experimental results showed that the side-view profile plus skin wrinkles could correctly differentiate between expressions, although they used acted rather than spontaneous facial expressions and had complex illumination conditions. Bando et al. [13] proposed a simple method for modelling wrinkles on human skin. They demonstrated that their method could model realistic wrinkle shapes by comparing them with a ground truth measure. However, despite providing a simple method to model wrinkles on human skin, this study lacked validation. Batool et al. [3] investigated forehead wrinkles, using a curve pattern as a soft biometric. Several metrics based on the Hausdorff distance and curve-to-curve correspondences were introduced to quantify the similarity of curves. However, information on curve position within a pattern was not included in the experiment.

In essence, wrinkle detection is a line or ridge detection problem. Only a few methods have been proposed in the literature. In texture analysis, Yap et al. [14] explored the multifractal approach detecting the key features of a human face. However, this method is not suitable for wrinkle detection as this approach finds repetitive patterns. Choi et al. [15] explored an accurate wrinkle representation scheme for skin age estimation. Their scheme showed that it may be possible to estimate face age automatically from the analysis of Local Binary Patterns (LBP). However, their work lacked validation of wrinkles being used. A popular class of approach to wrinkle detection are the snake-based methods, which use active contour maps to initialize and localize wrinkles [9]. Kwon and Lobo [9], [16] computed wrinkles from facial

C.-C. Ng, M.H. Yap and N. Costen are at the School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, Manchester, M1 5GD, UK. E-mail: {choon.c.ng, m.yap, n.costen}@mmu.ac.uk.

B. Li at the Department of Computer Science, Loughborough University, United Kingdom. Email: b.li@lboro.ac.uk.

Manuscript received xxxx xx, xxxx; revised xxxx xx, xxxx.

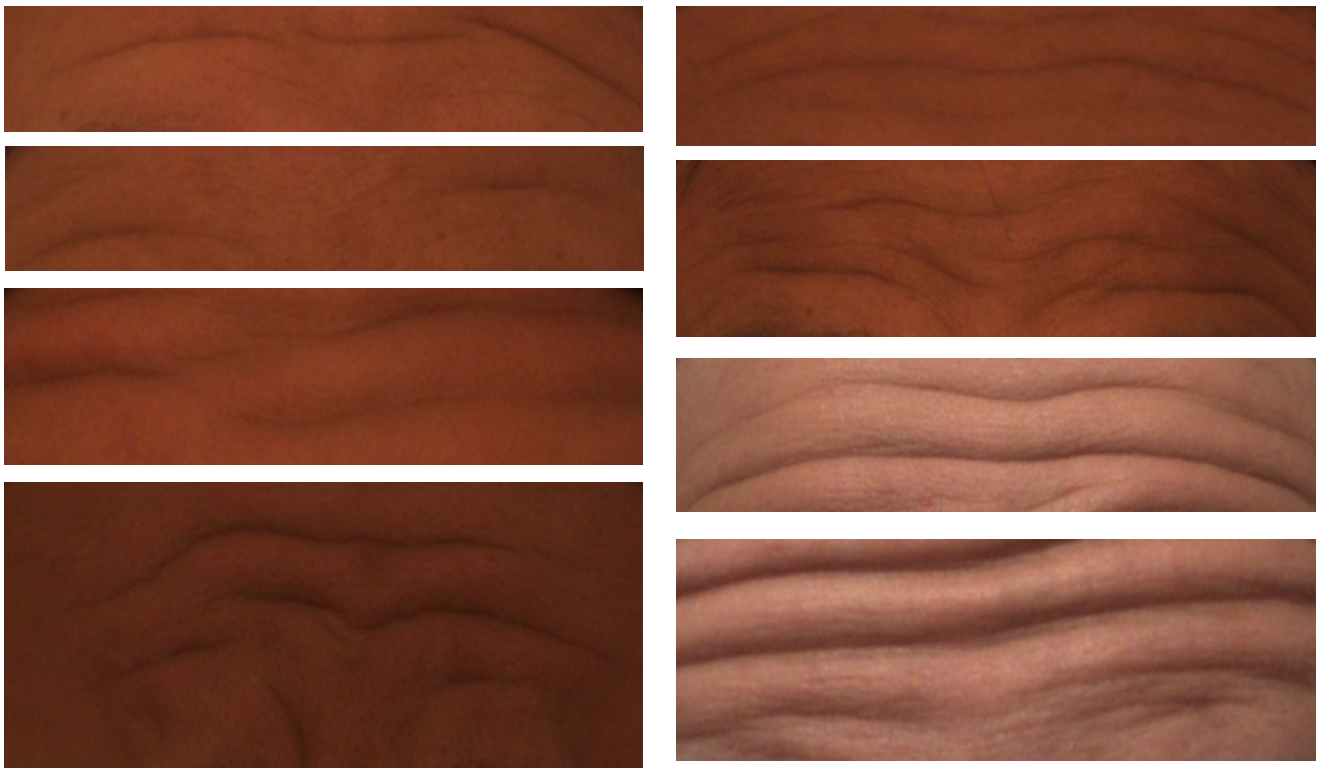


Fig. 1. Examples of the forehead dataset. Note that these images were resized for illustration purposes.

images to distinguish young and older adults. Wrinkles were computed in several regions, such as on the forehead, next to the eyes, and near the cheek bones. The presence of wrinkles in a region was based on the detection of curves in that area. However, the use of random initialization and multiple thresholds resulted in difficulties for automated detection. Alternatively, filter-based methods exploit local orientation, optimizing the response to ridge-like structures. Hayashi et al. [17] utilized an edge detection method and the Digital Template Hough Transform (DTHT) to extract both shorter and longer wrinkles on the face. This method could however fail because wrinkle detection is not necessarily a boundary detection problem.

Among the various wrinkle detection methods, that proposed by Cula et al. [7] (hereinafter referred as CUM) is the most representative. Cula et al. [7] explored the use of first-order derivative images and Gabor filtering for detecting wrinkle length and depth. The estimation was based on the orientation and frequency of elongated spatial features. The Wrinkle Index (WI) was defined as the product of both wrinkle length and wrinkle depth. Wrinkle length was the ratio between perimeter and area of the detected lines and wrinkle depth was derived from the Gabor filter responses. The authors claimed that the WI was highly correlated with the clinical scores and outputs of CUM. However, from a computer vision perspective, this work can be expanded by adding a detailed quantification of line segment in order to improve localization. A further validation of the accuracy of the wrinkle detection is also necessary.

Frangi et al. [18] proposed a vessel enhancement filter, known as the FRangi Filter (FRF), for extracting vessels from Magnetic Resonance Angiography (MRA) data on cerebral vasculature. The idea of FRF centres on the use of second-order partial derivatives for ridge detection. Eigenvalues of the Hessian matrix were used to extract principal directions into which the local second-order structure of the image can be decomposed. Since this gives the direction of the smallest curvature (along the vessel) directly, application of several filters with multiple orientations is avoided. Although FRF benefits from the presence of curve and valley extraction, a major drawback is its omni-directional nature. Both vertical and horizontal discontinuities are detected as wrinkles, however some of them are actually non-wrinkles.

Ng et al. [19] proposed a new method, referred to as a Hybrid Hessian Filter (HHF) to enhance the FRF. HHF is composed of both a directional gradient and a Hessian filter for wrinkle detection. It extracts important wrinkles by restricting the Hessian input to horizontal lines. This is an important step, as different parts of the face show different kinds of wrinkles. False detection of vertical lines appearing on the forehead, such as hairs and lighting effects could be minimized. Despite the fact that the HHF shows better accuracy than CUM or FRF, the relationship between the detected wrinkle and neighbourhood pixels has not yet been explored. The connectivity between wrinkle pixels is critical because a small area of wrinkles is always treated as noise and normally it will be removed in post-processing. Therefore, it is crucial to determine the gap between wrinkle pixels and potentially explore fine wrinkles in the surrounding area. In this study, a novel method, Hessian Line Tracking (HLT), is

BEGIN
$I(x,y) = ImageAcquisition()$
$(G_{\tilde{x}}, G_{\tilde{y}}) = ImageGradient(I)$
for $\sigma = FirstScale : step : LastScale$ **do**
  $H_{a,b,c} = GaussianConvolution(G_y, \sigma)$
  $\lambda_{1,2} = TextureOrientation(H_{a,b,c})$
  $\mathcal{R} = CurveDeviation(\lambda_{1,2})$
  $\mathcal{S} = SimilarityMeasure(\lambda_{1,2})$
  $\mathcal{E} = CurvelinearLikeliness(\mathcal{R}, \mathcal{S})$
end for
return $\arg\max_{\sigma} \{\mathcal{E}\}$
END

Fig. 2. Pseudo code of HHF.

proposed to identify the potential wrinkle pixels in the ridge area of the Hessian matrix.

The rest of this paper is organized as follows. Section II describes the image acquisition, summarises HHF and describes the proposed HLT. Then, Section III presents experimental results and discussion. In the final section, the outcome of this work is summarized.

II. Materials and Methods

A. Image Selection and Annotation

One hundred forehead images were cropped from the facial images of the Bosphorus dataset, which consists of varied sizes of forehead wrinkles [20]. This dataset consists of 106 participants, with the 2D face images acquired under good, flat illumination conditions with an ordinary camera. Each subject comprises several images of different poses and expressions. Only frontal image is used as the experimental dataset. Following Batool and Chellappa [3], generation of ground truth by hand labelling is most reliable, and wrinkles are more obvious on the forehead in most facial images. Therefore, we repeated their procedure by manually cropping the forehead sections from the Bosphorus dataset. Fig. 1 gives samples of the dataset used in this work.

Three coders were instructed to do the annotation in a controlled environment, using a custom-written Matlab application. After the three coders independently annotated each image judged to show wrinkles, the images were used for validation. For each image, only the deepest part of a wrinkle was annotated using a one-pixel line between its end points. The dimensions of the original images were 512 x 768 pixels. The cropped forehead images were different sizes due to the diversity of the facial structure. Resizing images was not undertaken because the manual cropping was considered sufficient to validate the reliability of the method in detecting wrinkles. Annotation was performed on colour images, but the subsequent image processing was carried out on grayscale images. The full acquisition procedure was described in Ng et al. [19]. Results of intra- and inter-reliability of manual annotation are discussed in the experimental section.

B. Hybrid Hessian Filter

BEGIN
$I = ImageAcquisition()$
$P = HybridHessianFilter(I)$
$sds = find(P > 0)$
for $s = InitScale : step : FinScale$ **do**
  for $t = 1 : length(sds)$ **do**
    $(x, y) = sds(t)$
    $[\mathbf{h}, \mathbf{p}, \mathbf{q}] = Neighborhood(x, y, s)$
    $V(\mathbf{h}) = CrossSectionalProfile(\mathbf{h}, \mathbf{p}, \mathbf{q})$
    $\theta = \arg\max_{\mathbf{h}} \{V(h_1) : V(h_6) > T\}$
    if $\theta != null$ **then**
      $D(x,y) = D(x,y) + 1$
      $(\hat{x}, \hat{y}) = (x + r\cos\theta, y + r\sin\theta)$
      while $(\hat{x}, \hat{y}) \notin sds$ **do**
        Repeat from step 9
      end while
    end if
  end for
end for
for all (i,j) in D **do**
  if $D(i,j) >= scaleSize$ **then**
    $wm(i,j) = 1$
  else
    $wm(i,j) = 0$
  end if
end for
$mf = MedianFiltering(wm)$
$df = DirectionalFiltering(mf)$
return $AreaThresholding(df)$
END

Fig. 3. Pseudo code of HLT.

Fig. 2 gives the pseudo code of the Ng et al. [19] HHF. The main idea is to approximate wrinkles based on a combination of Gaussian kernels and second order derivatives. Given a grayscale image $I(x,y)$, the directional gradient is computed, giving $(G_{\tilde{x}}, G_{\tilde{y}})$. Note that all calculations are computed at position (x,y). Since $G_{\tilde{y}}$ represents changes in the $\tilde{y}$ direction, it is suitable for horizontal line detection, and is used to calculate the Hessian matrix H at specific scale σ. Each of the approximations H_a, H_b, H_c is the result of convolution between $G_{\tilde{y}}$ and a Gaussian kernel. Then, the eigenvalues $\lambda_{1,2}$ of H are derived from each approximation in order to determine the texture orientation. After that, both curve deviation $\mathcal{R}$ and a similarity measure $\mathcal{S}$ are computed from $\lambda_{1,2}$. Then, the curvilinear likeliness $\mathcal{E}$ at specific σ is calculated from $\mathcal{R}$ and $\mathcal{S}$. Finally, the output is approximated from the maximum response of the filter centred on each pixel at a scale σ. For detailed HHF elaboration, please refer to Ng et. al. [19].

C. Proposed Hessian Line Tracking

HHF is capable of estimating the presence of the ridge and valley pattern typical of wrinkles. However, the relationship between wrinkle pixels with those surrounding them has not previously been explored. Therefore, the main focus of this paper is to introduce a new method, HLT, which has

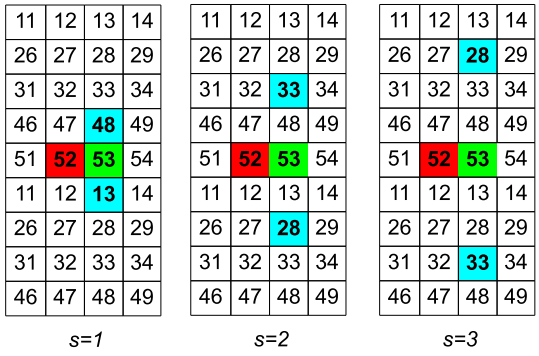


Fig. 4. Multi-scale tracking. Red is the current pixel, green is the candidate pixel and blue is the background pixel.

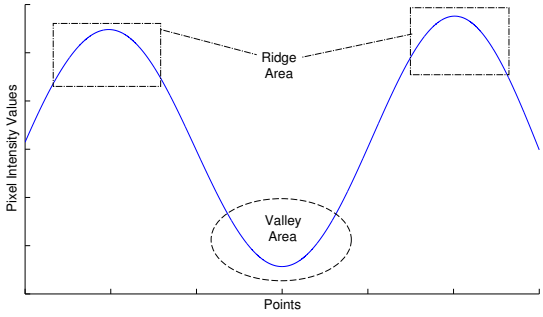


Fig. 5. Ideal wrinkle profile. Valley area presents the wrinkle while ridge area shows the non-wrinkle area.

the capacity to track wrinkle lines based on neighbourhood properties. HLT was inspired by a multi-scale retinal vessel segmentation technique, which uses line tracking, proposed by Vlachos and Dermatas [21]. However, they used the green colour channel to provide starting pixels or seeds for line tracking. This is not suitable in our case as the green channel is a weak representation of skin colour. The initialization step of seeding is crucial in HLT because too many seeds will lead to over-segmentation and too few seeds will cause under-segmentation. Therefore, the ridge area of the Hessian matrix was selected as the starting point of line tracking.

Fig. 3 depicts a pseudo-code version of the HLT. Several image processing modules are implemented: seed extraction by HHF provides the most appropriate pixels as starting points, line tracking at multiple scales determines the confidence of each pixel belonging to a wrinkle line, and post processing removes outliers. It is worth mentioning that HLT is applied to all seeds at multiple scales. The number of scales ($scaleSize$) is selected based on the size of wrinkles to be detected and also the resolution of the original image. The scales will increase if the wrinkle width is larger or the resolution is higher, and vice versa. Fig. 4 presents the multi-scale tracking of the same pixel with scales of 1, 2 and 3. In this study, we set the initial scale s to 1, in step of 1, and the last scale to 10.

First, a grayscale image I is selected from the dataset. The HHF is performed upon it, as described in the previous section and the output is defined as P. P is divided into ridge and valley areas. Fig. 5 illustrates an ideal wrinkle profile from non-wrinkle area to wrinkle area to non-wrinkle area. The

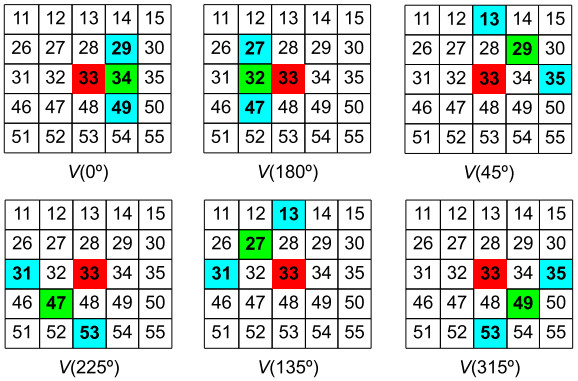


Fig. 6. Cross-sectional profile. Red is the current pixel, green is the candidate pixel and blue is the background pixel.

ridge is the area of interest as tracking always takes place from bright to dark areas. Therefore, any pixels of P larger than zero are selected as seeds sds. Let (x, y) denote the current pixel of sds, the connected pixels or candidate pixels $\mathbf{h}$ are neighbours of (x, y) that touch one of its edges or corners. Vertical pixels (top and bottom) are excluded because the tracking is only restricted to horizontal lines. Let $\mathbf{p}$ denote the first background pixel located s pixel(s) away from the candidate pixel and $\mathbf{q}$ a background pixel located s pixel(s) away from the candidate pixel but in the opposite direction. Fig. 6 illustrates the cross-sectional profile. The cross-sectional profile parameter V is estimated as

$$V(\mathbf{h}) = I(\mathbf{p}) + I(\mathbf{q}) - I(x, y) - I(\mathbf{h}) \tag{1}$$

If the current pixel belongs to the set of wrinkles, the cross-sectional parameter V has a large positive value. If the current pixel belongs to the background, the contrast between background pixels, current pixels and candidate pixels have similar values, thus V has a negative value or is near to zero. The angle θ with the maximum positive cross-sectional profile, exceeding a predefined positive threshold T, is defined as

$$\theta = \arg\max_{\mathbf{h}} \{V(\mathbf{h}) > T\} \tag{2}$$

where θ could be one of the elements of $\{0°, 180°, 45°, 225°, 135°, 315°, null\}$. T is an important threshold in identifying the intensity difference between the current pixel and the candidate pixel relative to the background pixels. The default value of T is 7. Vertical tracking is not included because the algorithm is only concerned with detecting horizontal wrinkles. If θ is not null, then the confidence array D of (x, y) is increased by one. If the candidate pixel is not drawn from the seeds, then the next contiguous pixel is tracked as

$$(\hat{x}, \hat{y}) = (x + r\cos\theta, y + r\sin\theta) \tag{3}$$

where r is the radius (default is unity) and θ is the angular coordinate. Tracking is repeated as shown in Fig 3. Otherwise, the calculation is continued by assigning a new current pixel to the next sds. Once multi-scale line tracking has been completed, the initial wrinkle map wm is generated by consulting

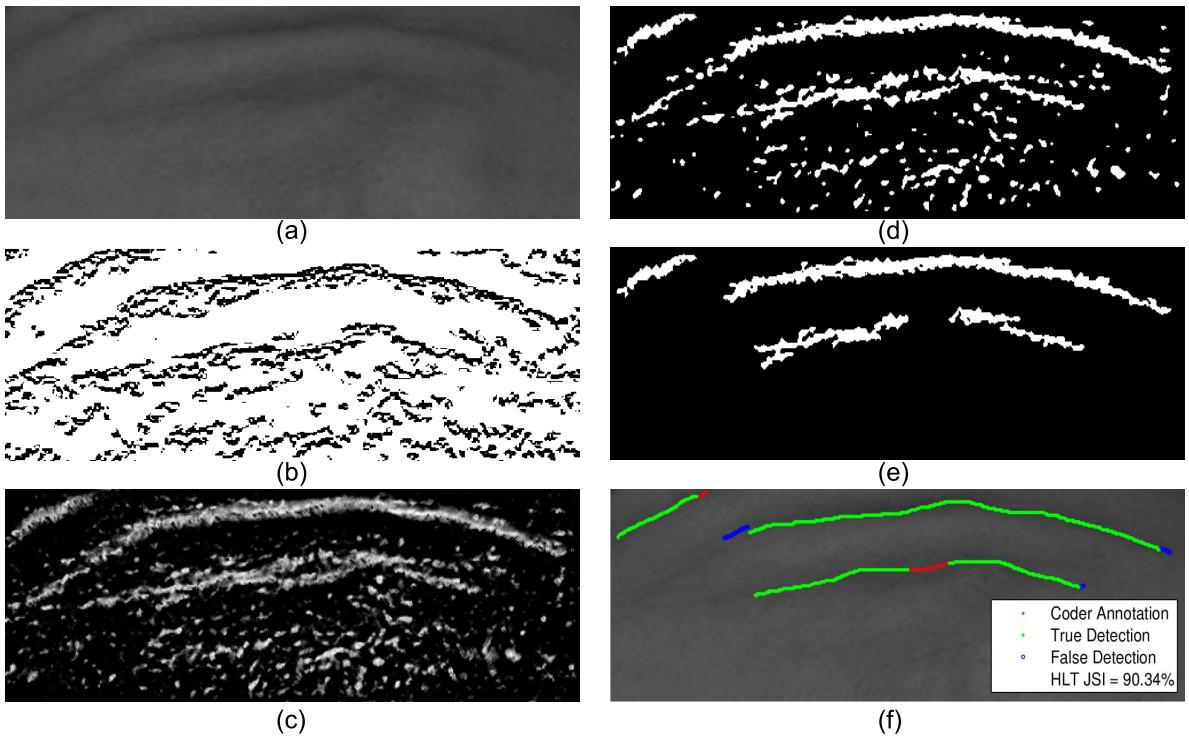


Fig. 7. HLT outputs of each processing step. (a) Original image I, (b) Seeds sds, (c) confidence array D, (d) initial wrinkle map w_m, (e) wrinkle map after post processing of filtering, (f) inter reliability results where red is ground truth, green is true positive and blue is false positive.

the confidence array. wm is *true* if $D(i,j)$ is greater than or equal to the number of scales. If the detection of wrinkles with a specific width is critical, wrinkles can be estimated from a predefined scale. In other words, the weights can be defined according to the width preference. For example, large weight values are preferred for coarse wrinkles and smaller for fine ones. In this study, the weight is equal to the number of scales which is expected to detect both coarse and fine wrinkles.

Fig. 7 shows the outputs of HLT, where (a), (b) and (c) illustrate the original image I, seeds sds and confidence array D, respectively. Due to noise, and very rarely, some non-wrinkle pixels can be tracked in all scales and are erroneously considered as part of the wrinkle map. As a final step, post processing is required to reduce the presence of such noisy pixels from the wrinkle map. This involves median filtering, directional filtering and area filtering. The median filtering is used to remove pepper noise pixels that do not belong to any wrinkles. A 3×3 median filter is applied to the wm and the output is mf. After median filtering, some noise may still remain. The binary image mf is transformed using four morphological openings with line structuring elements oriented in four different directions $0°$, $45°$, $90°$ and $135°$. These line structuring elements use a length of 3 pixels to preserve only wrinkle-like structures with lengths equal or larger than 3 pixels. df is derived using the logical OR of the four directional filtering responses and is shown as Fig. 7(d). Next, area thresholding is used to remove spots or roundish marks with a size is lower than 250 pixels. In the end, wrinkles are detected as illustrated in Fig. 7(e) and the center line of each wrinkle is used for validation as presented in Fig. 7(f).

D. Validation

In computer vision, we are interested on how closely a method can estimate a line at its desired location. In this context, we use the Jaccard Similarity Index (JSI) [22] to measure the reliability of the proposed wrinkle detection method. The Jaccard index,

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{4}$$

where A and B denote the set of pixels in two different coders' annotations.

In this study, an expansion of nine pixels is applied to the estimated line when calculating the JSI. We found by experimentation that this value produced the best overlapping in between the ground truth and estimated line. A small number of pixels will fail to hit the estimated line and a large number of pixels will bias the result as certain wrinkles are close to each other.

In order to validate the correctness of the proposed method, accuracy is defined as

$$accuracy = \sum_{i=1}^{N} w_i, \quad w_i = \begin{cases} 1 & \text{if } J_i > 40\% \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where N is the total number of images in the experiment and w is the logical output of J. Any overlap between A and B larger or equal to 40% is considered as correct detection [23].

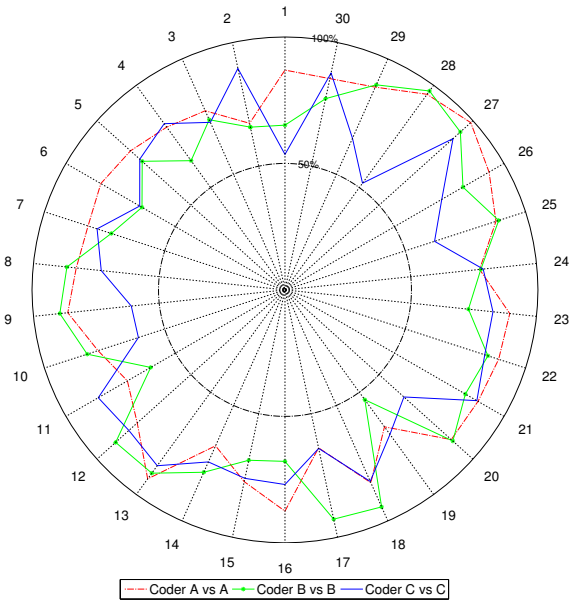


Fig. 8. Intra-reliability of the 30% dataset. Angle represents image number and radius illustrates the JSI percentage.

III. Results and Discussion

A. Reliability of Manual Annotation

In the first experiment, we validated the reliability of manual annotation by the coders. Coders were given the forehead images and they performed wrinkle annotation by saving the wrinkle mask in the Matlab logical format. Two of them were computer scientists and the third coder (denoted as C in the following experiments) was an experienced beauty therapist. The annotated masks were used to calculate the JSI within and between coders.

Fig. 8 shows the intra-reliability within coders. A round shape plot indicates the annotation is highly consistent. First the coders annotated the complete set of images on the first day, and then they were asked to repeat the annotation on 30% of the images on the second day. The intra-reliabilities of coders A, B and C in terms of average JSI are 82.55%, 78.99% and 74.56%, while the standard deviations are 8.58%, 11.28%, 10.45%, respectively. The intra-reliability of coder C is the lowest but the performance of coder B is the most inconsistent. From the manual annotation result, wrinkle annotation is not only a laborious and time-consuming task, but it is also a difficult perceptual task with low consistency within the same annotator.

In addition to intra-reliability, inter-reliability is a measurement of consistency among coders. Table I illustrates the JSI statistics between coders A, B and C. Overall, the accuracy of inter-reliability between coders is 94%. These results show that manual annotation is often subjective. Moreover, it is time consuming if manual annotation involves a large dataset. Therefore, an automatic wrinkle detection is needed to speed up and automate the annotation.

TABLE I
Inter-reliability of manual annotation. STD means standard deviation and Acc is accuracy.

Coders	Mean	STD	Min	Max	Acc
A vs B	67.28	14.10	18.41	94.88	99.00
A vs C	75.46	13.59	33.98	96.69	97.00
B vs C	61.46	16.92	23.05	98.34	86.00

B. Performance Assessment against Benchmark Algorithm

In this study, we repeated the HLT experiment based on the same dataset used by Ng et al. [19], but the annotation was improved by the inclusion of the beauty therapist. Fig. 9 presents the results of wrinkle detection using state-of-the-art methods CUM [7], FRF [18], HHF [19] and the proposed method HLT. Overall, we find that HLT performs better than CUM, FRF and HHF with an accuracy of 84% and STD 2.16%. The accuracies of CUM, FRF and HHF are 54.33%, 63.33% and 77.00% with STD 3.40%, 4.50% and 4.08%, respectively.

Fig. 10 shows some results of wrinkle detection using HHF and HLT. The first row is the original image, second row is the output of HHF and the last row is the HLT output. Fig. 10(a) shows that HHF achieved a JSI of 58.07% while HLT has 87.72%; Fig. 10(b) shows that HHF has a JSI of 40.07% while HLT has 63.59 %; Fig. 10(c) illustrates that HHF has a JSI of 3.54% while HLT has 50.28%; Fig. 10(d) depicts that HHF scores 34.34% while HLT is 87.45%.

A Student's t-test was performed between HHF and HLT, assessing the null hypothesis that there is no difference between HHF and HLT. This found that $t = 2.40$, $d.f. = 198$, $P = 0.02$, mean difference is 6.91, standard error difference is 2.88, upper of 95% confidence is 1.22 and lower is 12.60. As $P < 0.05$, the null hypothesis is rejected. In other words, there is a statistically significant difference between HHF and HLT.

Discussion

Wrinkles usually show a large variability in length, width and pattern in either different images and within the same image, challenging the development of automatic wrinkle detection. Therefore, an efficient multi-scale HLT has been proposed. This method is based on seed extraction by HHF and multi-scale line tracking to overcome for the weakness of HHF and capture wrinkle variability in the whole image. The number of seeds affects the robustness of the confidence array. If a small group of seeds is selected from the image, the detection accuracy of the algorithm will decrease because some wrinkle lines will not be reached from the seeds. On the other hand, the selection of an excessive number of seeds could increase the false positive rate. Therefore pre-processing must be undertaken in order to ensure that the selected group of seeds are representative, sufficient and minimal. HHF highlights potential wrinkle pixels based on a Gaussian distribution between ridge and valley curves. Ng et al. [19] found that wrinkle patterns were similar to the valley sequence. However, the ridge distribution in relation to wrinkles had not previously been explored. Therefore, in this study, ridge distribution is

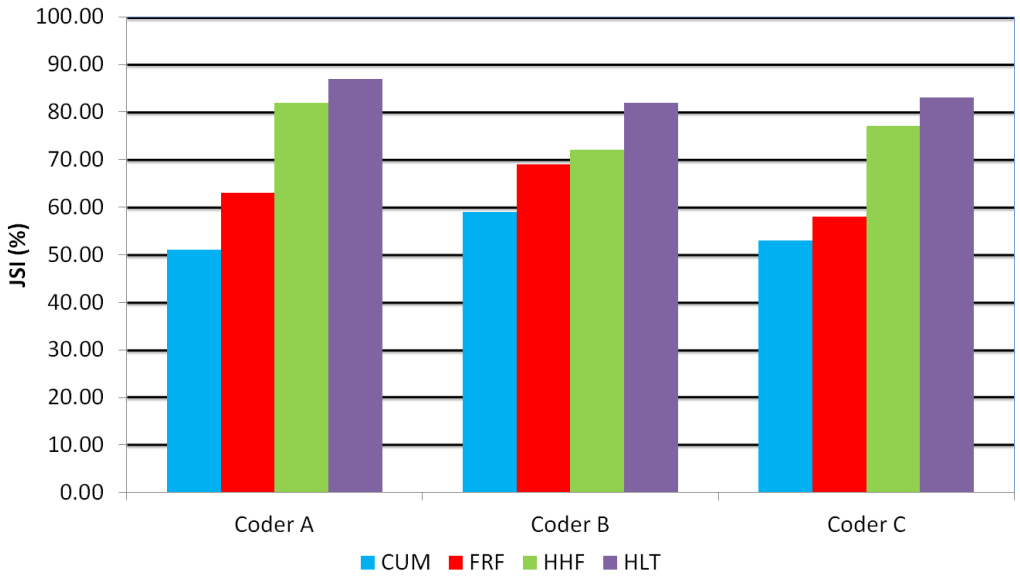


Fig. 9. JSI of automatic wrinkle detection versus benchmark of coder A, B and C.

used as the seed of line tracking and it produces significantly better results than the valley sequence. If tracking is limited to the valley sequence, the HLT detection is expected to be worse than HHF as the tracking is constrained by the valley region (darker pixels).

Originally, a line tracking method was proposed by Vlachos and Dermatas [21], but has some drawbacks if implemented it in the research context. First, the tracking involved all neighbourhood pixels. It computed the confidence array in all directions which is not ideal for directional wrinkles. If the target of interest is a horizontal wrinkle, the tracking should exclude vertical tracking. Therefore, both 90° and 270° in HLT are excluded from the tracking process with the assumption that the major orientation of forehead wrinkles is horizontal. In addition, the exclusion of the current or reference pixel in calculating the cross-sectional profile parameter V is not appropriate. In our opinion, it is important to show the relationship between the current pixel and candidate pixel to be tracked. If both have similar intensity and are darker than the background pixels, both pixels are highly similar and possibly belong to a wrinkle. Therefore, the calculation of the cross-sectional profile parameter has been modified by including the current pixel. Thirdly, threshold T is modified to adapt to the image noise. By default, the value is set to 7 for all cases. However, this value is not appropriate for all kinds of background. If specular light reflection is too strong, it will lead to over segmentation, otherwise, if the image is too smooth or the wrinkle to be detected is too fine, then under segmentation would happen. In order to reduce both problems, in HLT, an adaptive threshold is introduced. If the largest area of connected pixels in the initial wrinkle map is higher than 10% of the image size, then the threshold is set to a higher value of 15 to reduce the search space. If an empty map is generated, then the threshold is set to a lower value of 3. These parameters were tested in one dataset. For other datasets, it was necessary to adjust the parameters according to the resolution and lighting conditions.

Although the overall performance of HLT is better than HHF, there are a few issues to be addressed. Among 100 images of HHF, 18 performed worse than HLT. These images have a mixture of coarse and fine wrinkles. Moreover, three images with fine wrinkles give zero detection with either method. This issue could be caused by the nature of the image where the line is not very clear and even the coders achieved low inter-reliability on these images. In the future, additional computations such as wrinkle depth, length and width will be investigated. In addition, a curve fitting algorithm and shape modelling in relation to muscle movement could be explored, as this will benefit different fields such as age estimation [12] and micro movement detection [24].

IV. CONCLUSION

This paper proposed a new algorithm, Hessian Line Tracking, for wrinkle detection. It begins with a group of seeds, extracted from the ridge area of the Hessian matrix. Then, a multi-scale tracking system is applied recursively to all seeds. Once completed, each pixel confidence is validated over the scales to produce an initial wrinkle map. Finally, post-processing includes median filtering, directional filtering and area thresholding are applied for noise reduction. In the experimental setup, three coders were instructed to annotate the wrinkle manually on 100 cropped forehead images from Bosphorus dataset. As an assessment of manual annotation, both intra- and inter-reliability achieved accuracies of 94% and above, demonstrating that the manual annotation is reliable between and within coders. In the performance assessment against benchmark algorithms, HLT performs better than CUM, FRF and HHF with accuracies of 84.00%, 54.33%, 63.33% and 77.00%, respectively. Such performance is close to that of manual annotation. HLT's strengths includes the ability to explore wrinkle connectivity in addition to the curve and valley pattern.

ACKNOWLEDGMENT

This work was supported by a PhD Studentship from Manchester Metropolitan University. The authors would like to thank the coders for their valuable time in manual annotation.

REFERENCES

[1] J. Aznar-Casanova, N. Torro-Alves, and S. Fukusima, "How much older do you get when a wrinkle appears on your face? Modifying age estimates by number of wrinkles," *Aging, Neuropsychology, and Cognition*, vol. 17, no. 4, pp. 406–421, 2010.

[2] Y. Huang, Y. Li, and N. Fan, "Robust symbolic dual-view facial expression recognition with skin wrinkles: local versus global approach," *IEEE Trans. on Multimedia*, vol. 12, no. 6, pp. 536–543, 2010.

[3] N. Batool and R. Chellappa, "Detection and inpainting of facial wrinkles using texture orientation fields and Markov random field modeling," *IEEE Trans. On Image Processing*, vol. 23, no. 9, pp. 3773–3788, 2014.

[4] V. Gartstein and S. A. Shaya, "Image analysis of facial skin features," in *Application of Optical Instrumentation in Medicine XIV and Picture Archiving and Communication Systems (PACS IV) for Medical Applications*. International Society for Optics and Photonics, 1986, pp. 284–289.

[5] K. Mizukoshi and K. Takahashi, "Analysis of the skin surface and inner structure around pores on the face," *Skin Research and Technology*, vol. 20, no. 1, pp. 23–29, 2014.

[6] S. Luebberding, N. Krueger, and M. Kerscher, "Comparison of validated assessment scales and 3D digital fringe projection method to assess lifetime development of wrinkles in men," *Skin Research and Technology*, vol. 20, no. 1, pp. 30–36, 2014.

[7] G. O. Cula, P. R. Bargo, A. Nkengne, and N. Kollias, "Assessing facial wrinkles: automatic detection and quantification," *Skin Research and Technology*, vol. 19, no. 1, pp. e243–e251, 2013.

[8] K. Tsukahara, Y. Takema, H. Kazama, Y. Yorimoto, T. Fujimura, S. Moriwaki, T. Kitahara, M. Kawai, and G. Imokawa, "A photographic scale for the assessment of human facial wrinkles," *Journal of Cosmetic Science*, vol. 51, no. 2, pp. 127–139, 2000.

[9] Y. Kwon and N. da Vitoria Lobo, "Age classification from facial images," *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1–21, 1999.

[10] N. Ramanathan, R. Chellappa, and S. Biswas, "Computational methods for modeling facial aging: a survey," *Journal of Visual Languages & Computing*, vol. 20, no. 3, pp. 131–144, 2009.

[11] Y. Fu, G. Guo, and T. Huang, "Age synthesis and estimation via faces: a survey," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010.

[12] C.-C. Ng, M. H. Yap, N. Costen, and B. Li, "An investigation on local wrinkle-based extractor of age estimation," in *Proceedings of the 9th International Conference on Computer Vision Theory and Applications*, vol. 1, 2014, pp. 675–681.

[13] Y. Bando, T. Kuratate, and T. Nishita, "A simple method for modeling wrinkles on human skin," in *Proceedings of the 10th Pacific Conference on Computer Graphics and Applications*. IEEE, 2002, pp. 166–175.

[14] M. H. Yap, H. Ugail, R. Zwiggelaar, B. Rajoub, V. Doherty, S. Appleyard, and G. Hurdy, "A short review of methods for face detection and multifractal analysis," in *International Conference on CyberWorlds*. IEEE, 2009, pp. 231–236.

[15] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, "Age estimation using a hierarchical classifier based on global and local facial features," *Pattern Recognition*, vol. 44, no. 6, pp. 1262–1281, 2011.

[16] Y. Kwon and N. da Vitoria Lobo, "Age classification from facial images," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1994, pp. 762–767.

[17] J.-I. Hayashi, M. Yasumoto, H. Ito, and H. Koshimizu, "Age and gender estimation based on wrinkle texture and color of facial images," in *Int. Conf. on Pattern Recognition*, vol. 1. IEEE, 2002, pp. 405–408.

[18] A. F. Frangi, "Three-dimensional model-based analysis of vascular and cardiac images," PhD Thesis (Chapter 2), University Medical Center Utrecht, 2001.

[19] C.-C. Ng, M. Yap, N. Costen, and B. Li, "Automatic wrinkle detection using hybrid Hessian filter," in *The 12th Asian Conference on Computer Vision*. Springer International Publishing, 2014, pp. 609–622.

[20] A. Savran, B. Sankur, and M. Taha Bilge, "Regression-based intensity estimation of facial action units," *Image and Vision Computing*, vol. 30, no. 10, pp. 774–784, 2012.

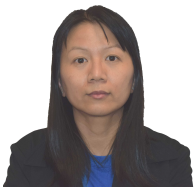
[21] M. Vlachos and E. Dermatas, "Multi-scale retinal vessel segmentation using line tracking," *Computerized Medical Imaging and Graphics*, vol. 34, no. 3, pp. 213–227, 2010.

[22] P. Jaccard, "Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines (in French)," *Bulletin de la Société Vaudoise des Sciences Naturelles*, vol. 37, no. 140, pp. 241–272, 1901.

[23] R. Real, "Tables of significant values of Jaccard's index of similarity," *Misc. Zool.*, vol. 22, no. 1, pp. 29–40, 1999.

[24] A. K. Davison, M. H. Yap, N. Costen, K. Tan, C. Lansley, and D. Leightley, "Micro-facial movements: An investigation on spatio-temporal descriptors," in *European Conference on Computer Vision*. Springer, 2014, pp. 111–123.

Choon-Ching Ng received B.Sc.(Hons.) and M.Sc. in Computer Science from Universiti Teknologi Malaysia in 2007 and 2010, respectively. From 2010 to 2012, he was a lecturer at University Malaysia Pahang. He is currently a PhD student attached to the School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University. His research interests include facial aging analysis, pattern recognition and natural language processing.



Moi Hoon Yap received her PhD in Computer Science from Loughborough University in 2009. She received her BSc (hons) in Statistics and MSc IT from Universiti Putra Malaysia. After her PhD, she worked as a Postdoctoral Research Assistant in the Centre for Visual Computing, University of Bradford. Currently, she is a Senior lecturer in Manchester Metropolitan University. Her research interests are facial analysis, medical image analysis, image and video processing.

Nicholas Costen received a B.A. in Experimental Psychology from the University of Oxford and Ph.D. degree in Mathematics and Psychology from the University of Aberdeen. He has undertaken research at the Advanced Telecommunications Research Laboratory, Kyoto and at the Division of Imaging Science and Biomedical Engineering, University of Manchester. He is a Reader in Cognitive Computer Vision at MMU where his interests include on face recognition and human motion analysis.

Baihua Li received B.Sc. and M.Sc. degrees in Electronic Engineering from Tianjin University, China and Ph.D. degree in Computer Science from Aberystwyth University, UK. She is a Lecturer in the Department of Computer Science, Loughborough University, United Kingdom. Her current research interests include computer vision, image processing, pattern recognition, advanced computer graphics, human motion analysis and behavior understanding from multi-modality imaging and sensory data.

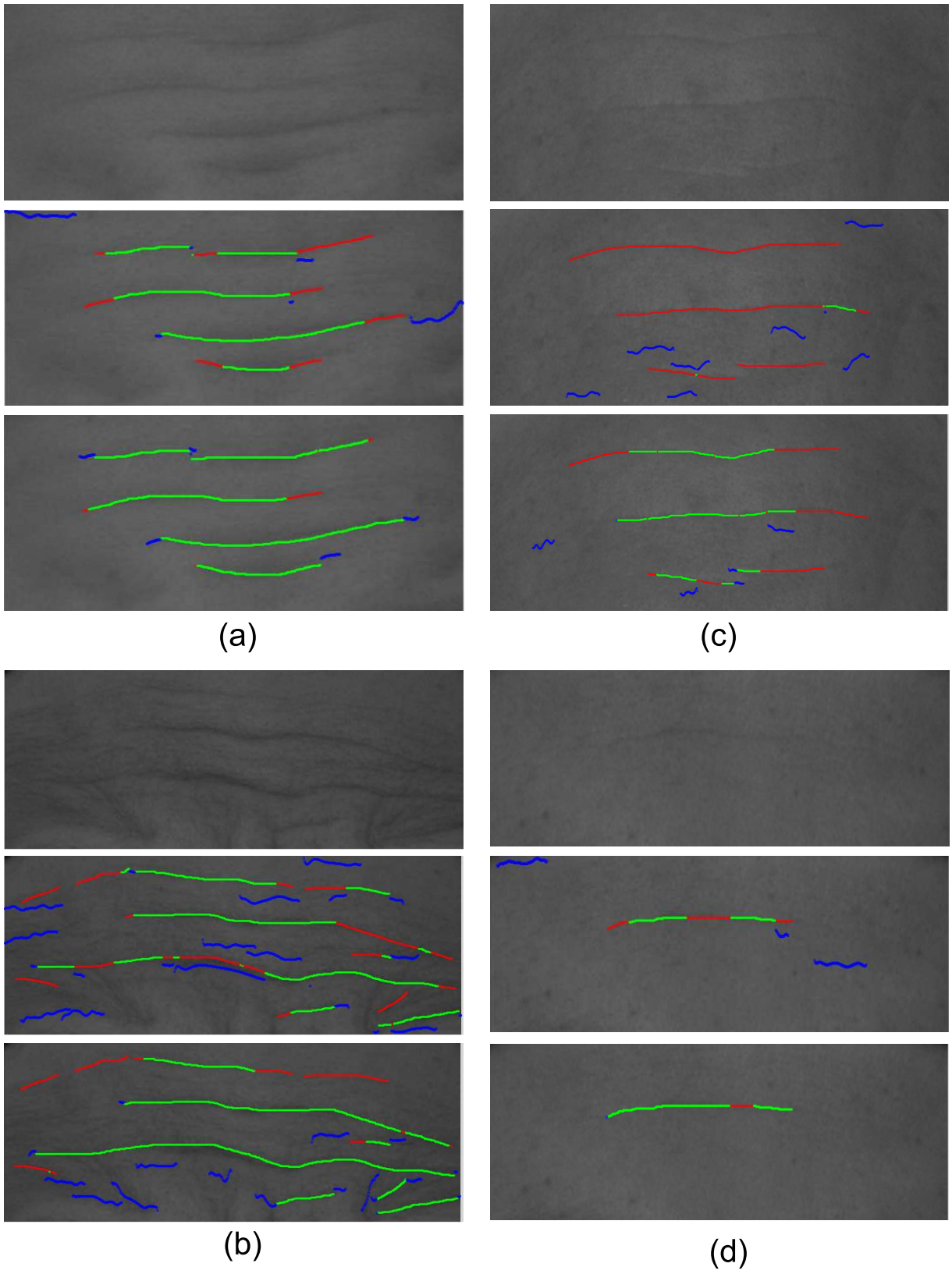


Fig. 10. HHF vs HLT of different images. Green is a hit on ground truth, red is a miss and blue is a false alarm. The first row is the original image, second row is the output of HHF and the last row is the HLT output. (a) JSI of HHF is 58.07% and HLT is 87.72%. (b) JSI of HHF is 40.07% and HLT is 63.59%. (c) JSI of HHF is 3.54% and HLT is 50.28%. (d) JSI of HHF is 34.34% and HLT is 87.45%.